

Limit distributions of upper order statistics for families of multivariate distributions

Mario V. Wüthrich

Received: 28 July 2004 / Revised: 3 May 2006 /
Accepted: 4 May 2006 / Published online: 12 July 2006
© Springer Science + Business Media, LLC 2006

Abstract We consider a portfolio of dependent exchangeable random variables X_1, \dots, X_n , where the dependence structure is generated by a mixture model (Archimedean copulas belong to this class of models). Define the ordered sample $X_{1,n} \geq X_{2,n} \geq \dots \geq X_{n,n}$. We prove results of the following type: fix $k \in \mathbb{N}$ and choose $(c_n, d_n)_{n \in \mathbb{N}}$ appropriately, then $(c_n^{-1}(X_{1,n} - d_n), \dots, c_n^{-1}(X_{k,n} - d_n))$ converges in distribution to a random vector $(Y^{(1)}, \dots, Y^{(k)})$ as $n \rightarrow \infty$, for which we can explicitly give the distribution.

Keywords Mixture models · Archimedean copula · Dependent random variables · Upper order statistics · Extreme value theory · Near-maximum

AMS 2000 Subject Classification Primary—62E20, 62H20, 62P05

1 Introduction and main result

The study of dependent sequences is a central topic in extreme value theory. In this work we study the asymptotic limit distribution of upper order statistics for a class of dependent exchangeable random variables.

One finds a wide list of references on the asymptotic behaviour of upper order statistics, e.g., results for i.i.d. sequences and stationary sequences can be found in Embrechts et al. (1997), Falk et al. (1994), Galambos (1987), Leadbetter et al. (1983), Reiss (1989), or Resnick (1987).

An excellent reference on the behaviour of upper order statistics for exchangeable random variables is Galambos (1987), Section 3.6. In this work we consider a multivariate mixture model which allows for an explicit calculation of limit distributions. The definition goes via Laplace transforms [see Marshall and Olkin (1988), formula (2.6)]: let F and M_Θ be univariate distributions with $M_\Theta(0) = 0$.

M. V. Wüthrich (✉)

Department of Mathematics, ETH Zürich, CH-8092 Zürich, Switzerland
e-mail: wueth@math.ethz.ch

Denote by ψ the Laplace transform of M_Θ , and by ψ^{-1} its inverse. Then we assume that X_1, \dots, X_n has the following distribution

$$G(x_1, \dots, x_n) = \int_0^\infty \prod_{i=1}^n \exp\{-\psi^{-1}(F(x_i))\}^\theta dM_\Theta(\theta). \quad (1.1)$$

Remarks:

- Formula (1.1) defines an n -dimensional random vector with exchangeable coordinates. Conditionally, given $\Theta = \theta$, they are i.i.d. And since an i.i.d. sequence can be extended to an infinite sequence, the vector X_1, \dots, X_n is in fact a segment of an infinite sequence of exchangeable random variables (see e.g., Galambos, 1987, Section 3.6).
- The Laplace transform is strictly decreasing with $\psi(0) = 1$ and $\psi(\infty) = 0$.
- Multivariate distributions of the form Eq. 1.1 can be represented with help of Archimedean copulas (see e.g., Marshall and Olkin, 1988).

Definition 1.1:

- For a random sequence X_1, X_2, \dots, X_n we define the ordered sample as follows

$$X_{1,n} \geq X_{2,n} \geq \dots \geq X_{n,n}, \quad (1.2)$$

hence $X_{k,n}$ is the k -th upper order statistic of the sequence X_1, \dots, X_n .

- An univariate distribution F belongs to the maximum domain of attraction of H , if there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that for all x : $F^n(c_n x + d_n) \rightarrow H(x)$ as $n \rightarrow \infty$. Write $F \in \text{MDA}(H)$.

Remark: For literature on maximum domains of attraction and limiting distributions H we refer to Chapter 3 in Embrechts et al. (1997).

Next we state our main result. We provide its proof in Section 3, below. The proof is a direct consequence of representation (1.1) and a theorem on limit distributions of upper order statistics for i.i.d. sequences of random variables (see e.g., Theorem 4.2.8 in Embrechts et al., 1997).

Theorem 1.2 (Joint limit distribution of k upper order statistic): We assume that X_1, \dots, X_n satisfies model assumptions (1.1) for all $n \geq 1$, moreover let $\exp\{-\psi^{-1} \circ F(\cdot)\} \in \text{MDA}(H)$ with norming constants $c_n > 0$ and $d_n \in \mathbb{R}$. Then for every $k \in \mathbb{N}$ we have

$$\left(\frac{X_{1,n} - d_n}{c_n}, \dots, \frac{X_{k,n} - d_n}{c_n} \right) \xrightarrow{d} (Y^{(1)}, \dots, Y^{(k)}), \text{ as } n \rightarrow \infty, \quad (1.3)$$

where $(Y^{(1)}, \dots, Y^{(k)})$ is a k -dimensional (ψ, H) -extremal variate with joint density: for $x_k < \dots < x_1$

$$h_{\psi,k}(x_1, \dots, x_k) = (-1)^k \psi^{(k)}(-\log H(x_k)) \prod_{i=1}^k \frac{h(x_i)}{H(x_i)}. \quad (1.4)$$

Remarks:

- For this specific form of exchangeable dependence structures we can explicitly determine the limit distributions of upper order statistics.
- For $\psi(x) = e^{-x}$ we have the independent case (see Embrechts et al., 1997, Definition 4.2.7), for $k = 1$ we obtain the distribution of the maximum (see Wüthrich, 2004, Theorem 3.2).

2 Application: on the number of near-maximum

Choose $a > 0$ fixed and introduce

$$K_n(a) = \sum_{i=1}^n 1_{\{X_i \geq X_{1,n} - a\}}, \quad (2.1)$$

this is the number of points that are within the random interval $[X_{1,n} - a, X_{1,n}]$. Various properties of $K_n(a)$ are proved by several authors under the i.i.d. assumption [see e.g., Pakes and Steutel (1997), Khmaladze et al. (1997), Li and Pakes (1998), Pakes and Li (1998)].

Applications to insurance were first considered in Li and Pakes (2001): let $N(t)$ be an integer valued point process independent of $\{X_k, k \geq 1\}$. Define

$$\mathcal{K}_t(a) = K_{N(t)}(a) = \sum_{i=1}^{N(t)} 1_{\{X_i \geq X_{1,N(t)} - a\}}, \quad (2.2)$$

then Li and Pakes (2001) study the asymptotic behaviour of \mathcal{K}_t for $N(t) \xrightarrow{P} \infty$ as $t \rightarrow \infty$ (in the i.i.d. case). These results are generalized in Hashorva (2003) to a sequence X_1, X_2, \dots of dependent random variables. We show that our model is an explicit example to Hashorva's results.

Corollary 2.1: Assume that X_1, X_2, \dots satisfies the assumptions of Theorem 1.2. Then we have for all $k \in \mathbb{N}$

$$\frac{X_{1,n} - X_{k,n}}{c_n} \xrightarrow{d} W_k^* = Y^{(1)} - Y^{(k)}, \quad (2.3)$$

where $(Y^{(1)}, Y^{(k)}) \sim h_{1,k}^\psi$ has density given for $x_k < x_1$ by

$$\begin{aligned} h_{1,k}^\psi(x_1, x_k) &= (-1)^k \psi^{(k)}(-\log H(x_k)) \frac{h(x_k)}{H(x_k)} \frac{h(x_1)}{H(x_1)} \\ &\quad \times \frac{[\log H(x_1) - \log H(x_k)]^{k-2}}{(k-2)!}. \end{aligned} \quad (2.4)$$

Corollary 2.2: Assume that X_1, X_2, \dots satisfies the assumptions of Theorem 1.2. Then for all $j \in \mathbb{N}$ and $\xi > 0$ we have

$$\lim_{n \rightarrow \infty} P[K_n(c_n \xi) < j] = P[W_j^* > \xi]. \quad (2.5)$$

Now we are ready to restate Corollary 2.4 of Hashorva (2003) for our situation:

Corollary 2.3: Assume that X_1, X_2, \dots satisfies the assumptions of Theorem 1.2. Let $N(\cdot)$ be an integer valued point process independent of $\{X_k, k \geq 1\}$. Suppose there exists a positive random variable Z with

$$\frac{N(t)}{t} \xrightarrow{P} Z \quad \text{for } t \rightarrow \infty. \quad (2.6)$$

Then we have

$$\lim_{t \rightarrow \infty} P[\mathcal{K}_t(c_{[t]}\xi) < j] = P[Z^\gamma W_j^* > \xi], \quad (2.7)$$

where γ is characterized by H : i) if $MDA(H)$ is the Gumbel domain then $\gamma = 0$; ii) if $MDA(H)$ is the Fréchet domain then $\gamma = 1/\alpha$; iii) if $MDA(H)$ is the Weibull domain then $\gamma = -1/\alpha$ (where α is defined in Theorem 3.2.3 of Embrechts et al., 1997). Moreover Z and W_j^* are independent.

Remark (Convergence in distribution): Observe that if we would have that $W_j^* \xrightarrow{P} \infty$, respectively, $Z^\gamma W_j^* \xrightarrow{P} \infty$, (as $j \rightarrow \infty$) we would know that there exist random variables \mathcal{W}_ξ^* and \mathcal{W}_ξ such that

$$K_n(c_n\xi) \xrightarrow{d} \mathcal{W}_\xi^*, \quad \text{respectively } \mathcal{K}_t(c_t\xi) \xrightarrow{d} \mathcal{W}_\xi, \quad (2.8)$$

with $P[\mathcal{W}_\xi^* < j] = P[W_j^* > \xi]$ and $P[\mathcal{W}_\xi < j] = P[Z^\gamma W_j^* > \xi]$. But examples in Hashorva (2003) show that this is not necessarily the case.

Example 2.4: [Section 5.2 Wüthrich, 2004, Fréchet case revisited]

We choose the following model: we assume that X_1, X_2, \dots satisfy Eq. 1.1 with Pareto marginals and Gumbel copula, i.e., for $K, \beta > 0$, $\gamma \geq 1$

$$1 - F(x) \sim Kx^{-\beta} \quad \text{as } x \rightarrow \infty, \quad (2.9)$$

$$\psi^{-1}(t) = (-\log t)^\gamma. \quad (2.10)$$

Hence we are in the Fréchet domain with $\alpha = \beta \cdot \gamma$. [see Wüthrich (2004), (5.11)–(5.12)]. Moreover we know that the maximum $X_{1,n}$ grows with speed $n^{1/\beta\gamma}$ as $n \rightarrow \infty$. Hence we easily see that with increasing dependence strength γ the speed of growth of $X_{1,n}$ decreases (the stronger the dependence the harder it is for a single coordinate to escape). Therefore, on the other hand, we know that the j -th largest coordinate $X_{j,n}$ lies within distance of order $n^{1/\beta\gamma}$ of the maximum $X_{1,n}$.

Concluding remarks. We have obtained explicit asymptotic distributions for upper order statistics satisfying model assumptions (1.1). Unlike in most other applications our model does not really reflect time series of random variables. Our model is rather motivated by considering large non-life insurance contracts with identical dependent risks. The latent variable Θ reflects the risk characteristics of different accident years. It measures the effects within the portfolio, that do not diversify. This consideration is very close to Bayesian and credibility ideas.

3 Proofs

Proof of Theorem 1.2: Choose c_n, d_n such that $\exp\{-\psi^{-1} \circ F\} \in MDA(H)$. For $x_1 \geq x_2 \geq \dots \geq x_k$ we define $u_n^{(i)} = c_n x_i + d_n$. Hence, we find for $k \geq n$

$$P[X_{1,n} \leq u_n^{(1)}, \dots, X_{k,n} \leq u_n^{(k)}] = E[P[X_{1,n} \leq u_n^{(1)}, \dots, X_{k,n} \leq u_n^{(k)} | \Theta]], \quad (3.1)$$

where conditionally, given Θ , (X_1, \dots, X_n) are independent with marginal distributions $\exp\{-\psi^{-1} \circ F\}^\Theta$. Hence, $\exp\{-\psi^{-1} \circ F\}^\Theta \in MDA(H^\Theta)$. This implies that conditionally, given Θ , X_1, X_2, \dots satisfies the classical Fisher–Tippett theorem with normalizing constants c_n, d_n and limiting distribution $H(\cdot)^\Theta$ (see Embrechts et al., 1997, Theorem 3.2.3, and Galambos, 1987, Section 3.6). Henceforth, for the upper order statistic $(X_{1,n}, \dots, X_{k,n})$, given Θ , we apply Theorem 4.2.8 of Embrechts et al. (1997), which gives

$$\begin{aligned} & \lim_{n \rightarrow \infty} P[X_{1,n} \leq u_n^{(1)}, \dots, X_{k,n} \leq u_n^{(k)} | \Theta] \\ &= \int_{-\infty}^{x_k} dy_k \int_{y_k}^{x_{k-1}} dy_{k-1} \cdots \int_{y_2}^{x_1} dy_1 H^\Theta(y_k) \prod_{i=1}^k \frac{\Theta h(y_i)}{H(y_i)}, \end{aligned} \quad (3.2)$$

i.e., we have pointwise convergence in Θ . Moreover, the probability is uniformly bounded by 1, which is integrable on $[0, \infty)$ with respect to $dM_\Theta(\theta)$. Hence we can apply the dominated convergence theorem, which implies

$$\begin{aligned} & \lim_{n \rightarrow \infty} P[X_{1,n} \leq u_n^{(1)}, \dots, X_{k,n} \leq u_n^{(k)}] \\ &= \int_{-\infty}^{x_k} dy_k \int_{y_k}^{x_{k-1}} dy_{k-1} \cdots \int_{y_2}^{x_1} dy_1 \int_0^\infty dM_\Theta(\theta) H^\theta(y_k) \prod_{i=1}^k \frac{\theta h(y_i)}{H(y_i)} \\ &= \int_{-\infty}^{x_k} dy_k \cdots \int_{y_2}^{x_1} dy_1 \prod_{i=1}^k \frac{h(y_i)}{H(y_i)} (-1)^k \int_0^\infty dM_\Theta(\theta) (-\theta)^k \exp\{-\theta(-\log H(y_k))\}. \end{aligned} \quad (3.3)$$

Now we use that ψ is the Laplace transform of M_Θ . Hence the k -th derivative is given by

$$\psi^{(k)}(\lambda) = \frac{d^k}{d\lambda^k} \psi(\lambda) = \int dM_\Theta(\theta) (-\theta)^k \exp\{-\lambda\theta\}. \quad (3.4)$$

There remains to proof that we have total measure 1 for $(x_1, \dots, x_k) \rightarrow (\infty, \dots, \infty)$. We have [in the last step we use Theorem 3.2 of Wüthrich (2004) for $x_k \in \mathbb{R}$]

$$\begin{aligned} & P[X_{1,n} \leq u_n^{(1)}, \dots, X_{k,n} \leq u_n^{(k)}] \geq P[X_{1,n} \leq u_n^{(k)}, \dots, X_{k,n} \leq u_n^{(k)}] \\ &= P[X_{1,n} \leq u_n^{(k)}] \rightarrow \psi(-\log H(x_k)), \text{ as } n \rightarrow \infty. \end{aligned} \quad (3.5)$$

This last term goes to 1 as $x_k \rightarrow \infty$. This together with Eq. 3.3 finishes the proof of Theorem 1.2. \blacksquare

Proof of Corollary 2.1: The proof easily follows from Theorem 1.2 and a consideration similar to the proof of Proposition 2.5 in Hashorva (2003). ■

Proof of Corollaries 2.2–2.3: The proof of Corollary 2.2 follows from Corollary 2.1 and for $1 \leq j \leq n$ and $\xi > 0$

$$P[K_n(c_n\xi) < j] = P\left[\frac{X_{1,n} - X_{j,n}}{c_n} > \xi\right]. \quad (3.6)$$

The proof of Corollary 2.3 then follows from Propositions 2.1–2.2 and Corollary 2.4 in Hashorva (2003). ■

References

- Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin Heidelberg New York (1997)
- Falk, M., Hüsler, J., Reiss, R.D.: *Laws of Small Numbers: Extremes and Rare Events*. DMV Seminar Vol. 23. Birkhäuser, Basel (1994)
- Galambos, J.: *The Asymptotic Theory of Extreme Order Statistics*. 2nd edition. Krieger, Malabar (1987)
- Hashorva, E.: On the number of near-maximum insurance claim under dependence. *Insur. Math. Econ.* **32**, 37–49 (2003)
- Khmaladze, E., Nadareishvili, M., Nikabadze, A.: Asymptotic behaviour of a number of repeated records. *Statist. Probab. Lett.* **35**(1), 49–58 (1997)
- Leadbetter, M.R., Lindgren, G., Rootzén, H.: *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin Heidelberg New York (1983)
- Li, Y., Pakes, A.G.: On the number of near-records after the maximum observation in a continuous sample. *Comm. Statist. Theory Methods* **27**(3), 673–686 (1998)
- Li, Y., Pakes, A.G.: On the number of near-maximum insurance claims. *Insur. Math. Econ.* **28**(3), 309–323 (2001)
- Marshall, A.W., Olkin, I.: Families of multivariate distributions. *J. Am. Stat. Assoc.* **83**, 834–841 (1988)
- Pakes, A.G., Li, Y.: Limit laws for the number of near maxima via the Poisson approximation. *Statist. Probab. Lett.* **40**(4), 395–401 (1998)
- Pakes, A.G., Steutel, F.W.: On the number of records near the maximum. *Aust. J. Stat.* **39**, 172–192 (1997)
- Reiss, R.D.: *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer, Berlin Heidelberg New York (1989)
- Resnick, S.I.: *Extreme Values, Regular Variation and Point Processes*. Springer, Berlin Heidelberg New York (1987)
- Wüthrich, M.V.: Extreme value theory and Archimedean copulas. *Scand. Actuarial J.* 211–228 (2004)